Exploring Pharmacoepidemiologic Groupings of Drugs from a Clinical Perspective

Rainer Winnenburg, Olivier Bodenreider

National Library of Medicine, Maryland, MD, USA

Abstract

Objectives: To investigate the extent to which pharmacoepidemiologic groupings are homogeneous in terms of clinical properties. Methods: In our analysis, we classified drug subgroups from the pharmacoepidemiologic Anatomical Therapeutic Chemical (ATC) classification system based on clinical drug properties. We established mappings from ATC fifth level drug entities to drug property annotations in the National Drug File Reference Terminology (NDF-RT), including therapeutic categories, mechanisms of action, and physiologic effects. Based on the annotations for the individual drugs we computed homogeneity scores for all ATC groups and analyzed their distribution. Conclusions: We found ATC groups to be generally homogeneous, more so for mechanisms of action, and physiologic effects than for therapeutic intent. However, only half of all ATC drugs can be analyzed with this approach, in part because of missing properties in NDF-RT.

Keywords:

ATC, NDF-RT, pharmacoepidemiologic groups, clinical use.

Introduction

The Anatomical Therapeutic Chemical (ATC) classification system is a drug classification system developed by the World health Organization (WHO) Collaborating Centre for Drug Statistics Methodology in Norway. ATC groups drugs "according to the organ or system on which they act and their therapeutic, pharmacological and chemical properties." [1].

Traditional uses of ATC have included pharmacoepidemiologic studies. For example, ATC has supported the analysis of prescription drug costs [2] and use of drugs in various populations [3-6]. More recently, ATC groups have been related to various drug properties, including chemical similarity [7], drug-drug interactions [8], and adverse events in the EU-ADR project [9]. While ATC groupings have been shown to be adequate for pharmacoepidemiologic studies, it remains to be seen if they also support clinical applications, such as clinical decision support (CDS). For example, drug-drug interactions are often asserted between one drug and one drug group (e.g., between lithium and ACE inhibitors). Such CDS rules assume homogeneity of the drug groups, i.e., that the drugs from these groups have the same characteristics (e.g., therapeutic use, mechanisms of action and physiologic effects). Here the ATC group ACE inhibitors is homogeneous, because all its drugs (e.g., captopril, enalapril, lisinopril) share the same mechanism of action, namely 'Angiotensin-converting Enzyme Inhibitors'. However, ATC does not necessarily guarantee such homogeneity. In fact the ATC documentation specifically states that "Substances classified in the

same ATC fourth level cannot be considered pharmacotherapeutically equivalent since their mode of action, therapeutic effect, drug interactions and adverse drug reaction profile may differ."

The objective of this investigation is to explore ATC groupings from a clinical perspective. More specifically, we investigate the degree to which the ATC groupings are homogeneous with respect to properties such as therapeutic intent, mechanisms of action and physiologic effects. In practice, we use drug properties asserted in the National Drug File-Reference Terminology (NDF-RT) to explore the homogeneity of the ATC groupings.

While NDF-RT has been explored from various perspectives [10, 11], to our knowledge, this study is the first one to relate ATC groupings to drug properties in NDF-RT.

Materials

Anatomical Therapeutic Chemical (ATC) classification

As shown in Table 1, the ATC classification is organized into five levels. Ingredients are found at the fifth level, whereas levels one to four correspond to groupings. For the purpose of this study, we focus on the second and fourth levels. The second level roughly corresponds to therapeutic intent, while we expect the fourth level to reflect mechanisms of action and physiologic effects.

ATC comprises 90 subgroups at the second level, 780 subgroups at the fourth level, and 4464 drugs (fifth level). Not all the drugs in ATC are available on the U.S. market and the mapping of ATC drugs to U.S. drug resources is thus expected to be incomplete.

In this study, we use the 2012 version of ATC.

Table 1 – ATC hierarchy with examples

Code	Label	Level
C	Cardiovascular system	1 - anatomical
C01	Cardiac therapy	2 - therapeutic
C01A	Cardiac glycosides	3 - pharmacological
C01AA	Digitalis glycosides	4 - pharmaceutic
C01AA05	Digoxin	5 - drug

National Drug File-Reference Terminology (NDF-RT)

The National Drug File Reference Terminology (NDF-RT) is a resource developed by the Department of Veterans Affairs (VA) Veterans Health Administration, as an extension of the VA National Drug File [12]. The version used in this study is dated November 5, 2012. This version covers 7162 active moieties (level = ingredient). We used the NDF-RT API [13]

for mapping drug names to NDF-RT, as well as for querying drug properties.

NDF-RT is organized into several hierarchies. Across hierarchies, entities are related by various kinds of relations. For example, ingredients from the drug hierarchy are linked to entities from the mechanisms of action hierarchy by the relationship has_MoA (e.g., *ATORVASTATIN* has_MoA

Hydroxymethylglutaryl-CoA Reductase Inhibitors), to entities from the physiologic effect hierarchy by the relationship has_PE (e.g., *ATORVASTATIN* has_PE *Decreased Cholesterol Synthesis*). Therapeutic intent is expressed through relations between ingredients and entities from the disease hierarchy, including may_treat (*ATORVASTATIN* may_treat *Hypercholesterolemia*) and may_prevent (*ATORVASTATIN* may_prevent *Coronary Artery Disease*).
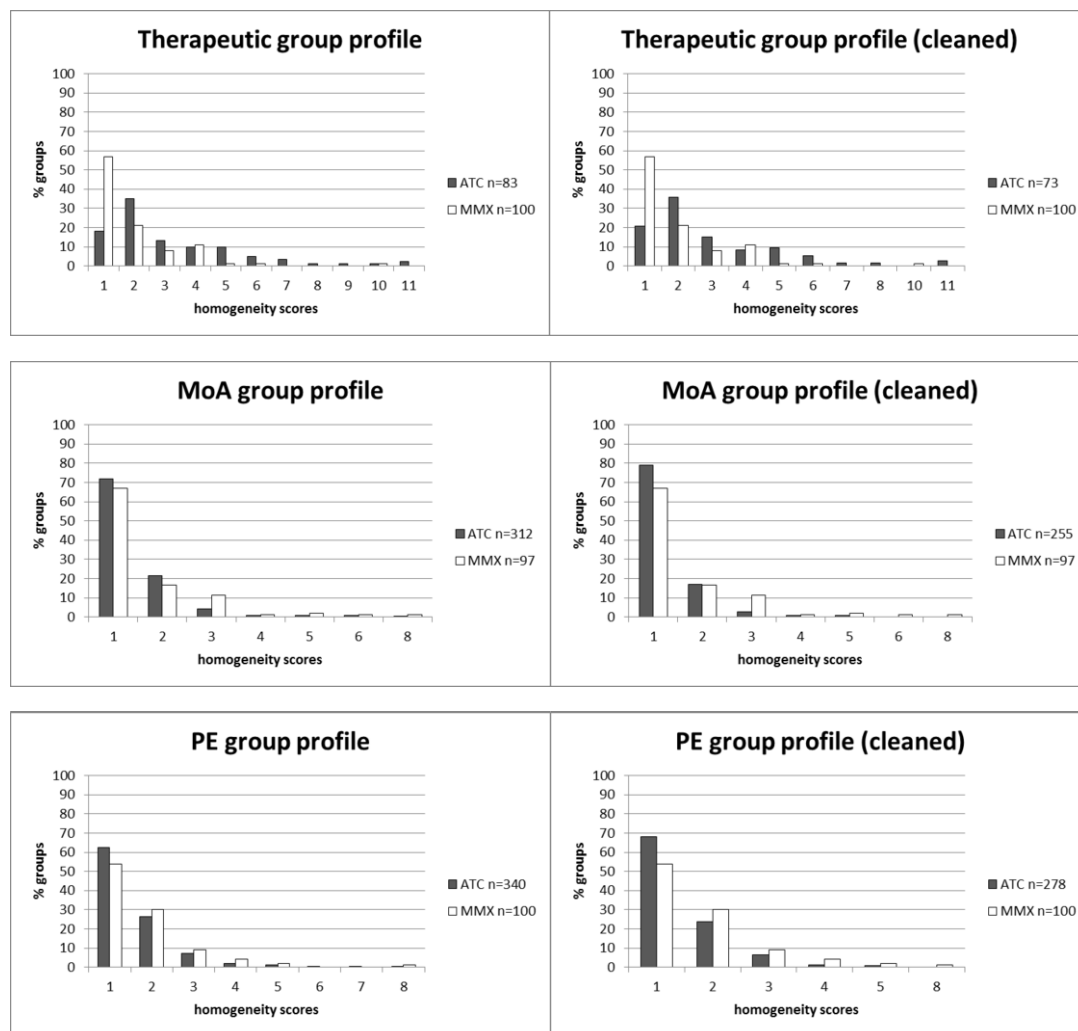

Figure 1 – Group profiles based on homogeneity scores for drug groups in ATC and classes in MMX. Homogeneity scores for thera-peutic intent are calculated for ATC second level groups, those for mechanisms of action (MoA) and physiologic effects (PE) on ATC fourth level groups. On the left-hand side, profiles include all groups, while on the right-hand side residual groups were ignored.

Methods

We designed methods for characterizing drug groups by clinical properties derived from the National Drug File Reference Terminology (NDF-RT). First, we mapped ATC drugs to NDF-RT drug concepts. Then we acquired the clinical properties from associations with the drug concepts present in NDF-RT. Based on the clinical properties of individual drugs, we computed homogeneity scores for each ATC second and fourth level subgroup. Finally, we studied the distribution of

homogeneity scores for all ATC groups at a given level ("pro-file") and compared ATC profiles to a clinical reference.

Mapping ATC drugs to NDF-RT clinical properties

Exclusions

Most of the ATC drug entities correspond to ingredients in NDF-RT, e.g., *Digoxin*, C01AA05 in ATC and N0000146388 in NDF-RT. However, some of the drugs in ATC have unspecific, collective terms (such as *thyroid gland preparations*) and others are out of scope for NDF-RT, such as radiopharmaceuticals. In the following we list the main categories of ATC drugs we considered out of scope, with examples:

- Multi-ingredient drugs (labels containing "*and*", "*other*", "*combinations*")
- Isotopes, especially group V, Radiopharmaceuticals (e.g., *iobenguane (131I)*)
- Unspecific, collective terms (e.g., *barium sulfate without suspending agents*)

On the other hand, in order to improve recall, we ignored extraneous information from the original ATC drug names, in the form of parenthetical expressions and appositions. For example, we used *nicotinyl alcohol* to map *nicotinyl alcohol (pyridylcarbinol)* to NDF-RT. Typical examples are:

- Terms with synonyms in parentheses (*nicotinyl alcohol (pyridylcarbinol)*)
- Overspecified terms (*immunoglobulins, normal human, for extravascular adm.)*

In total, we excluded 900 multi-ingredient drugs, drug combinations, and drugs out of scope from our mapping to NDF-RT.

Direct Mapping to NDF-RT

We used the NDF-RT API [13] to map the names of the ATC drugs (fifth level) to drug concepts at the ingredient level in NDF-RT (restricted to NDF-RT entities of kind *DRUG_KIND* and level *Ingredient*).

Indirect mapping via RxNorm

If an ATC drug name failed to be associated with clinical properties in NDF-RT, we assumed that the clinical properties were possibly associated with a related ingredient (i.e., non-salt ingredient vs. salt ingredient, e.g., *fluocinolone* vs. *fluocinolone acetonide*). In order to find such related ingredients, we used the RxNorm API [14] to map those drugs to their corresponding salt and non-salt ingredients, for which we then queried the clinical properties in NDF-RT as described above.

Acquiring clinical properties for drugs

For each drug, we extracted the list of clinical properties in the following three categories: 1) *therapeutic intent* (roles: *may_treat* and *may_prevent*), 2) *mechanism of action* (role: *has_MoA*) and 3) *physiologic effect* (role: *has_PE*). We ignored unspecific annotations containing the term *unknown*. In practice, we extracted the clinical properties for a given drug concept using the NDF-RT API method *getRelatedConceptsByRole*. Below is the list of clinical properties for the drug *cetirizine*.

Table 2 – List of clinical properties for the drug cetirizine

Role	Value
may_treat	Rhinitis, Allergic, Perennial
may_treat	Urticaria
has_MoA	Histamine H1 Receptor Antagonists
has_PE	Decreased Histamine Activity

Computing homogeneity scores

In the following, we explain how we computed the homogeneity scores that form the basis for the drug group profiles. For a given ATC subgroup, we aggregated all the clinical properties for all the drugs in that subgroup. More specifically, we collected all the therapeutic intent properties for each second-level subgroup. Similarly, we collected all the mechanism of action and physiologic effect properties, separately, for each fourth-level subgroup.

In order to assess the homogeneity of subgroups with respect to clinical properties, we counted how many distinct properties (or sets of properties) are necessary to account for at least 90% of the drugs in a given subgroup. For example, the fourth level subgroup R01AD *Corticosteroids* contains 10 drugs for which we could find mechanism of action properties in NDF-RT. These 10 drugs have the same mechanism of action: *Glucocorticoid Receptor Agonists*. Therefore, one single property accounts for 100% of the drugs in the subgroup and the homogeneity score for this subgroup is 1, which denotes the maximal homogeneity. In contrast, the fourth level subgroup V03AB *Antidotes* contains 12 drugs with annotations for mechanisms of action. The homogeneity score for this subgroup is 8, as eight distinct properties are needed to describe 90% of the drugs in the subgroup. The properties (or sets of properties) are as follows: *Siderophore Iron Chelating Activity, Cholinesterase Inhibitors, GABA B Antagonists, Free Radical Scavenging Activity, Alcohol Dehydrogenase Inhibitors,* {*Noncompetitive Opioid Antagonists, Competitive Opioid Antagonists*}, {*Adrenergic alpha1-Antagonists, Adrenergic alpha2-Antagonists*}, *and Cholinesterase Reactivators.*

Because the homogeneity score would be 1 mechanically for all groups containing only one drug, we only considered ATC drug groups that contain more than one drug.

We compiled the homogeneity scores for second level groups in ATC (with respect to therapeutic intent), and for fourth level subgroups (with respect to mechanism of action and physiologic effect, respectively).

We then created a profile of ATC groups corresponding to the distribution of homogeneity scores for all groups at a given level for the corresponding clinical property. In practice, we have three such profiles: for the second-level groups with the therapeutic intent property, and for the fourth-level groups with the mechanism of action and physiologic effect, respectively.

Comparison to the clinical reference Micromedex

We applied the same methods for concept mapping and computation of homogeneity scores to data we extracted from the clinical reference Micromedex. The Micromedex classes were extracted from a drug-drug interaction system for cases where the interaction was stated at the class level, but for which the list of drugs in a class was provided. Based on the assumption that a clinical reference should classify drugs more homogeneously with respect to clinical properties, we compared the distributions of homogeneity scores for all drug groups in ATC and classes in Micromedex.

Results

Mapping ATC drugs to NDF-RT clinical properties

Of the 4,464 drugs in ATC, 3,564 are single ingredient drugs excluding radiopharmaceuticals (see *Table 5*, column *Single ingredient drugs within scope),* of which we were able to map 2,111 (59%) to NDF-RT concepts (second column from the right). However, only 1,701 (48%) ATC drugs were mapped to NDF-RT ingredients associated with at least one clinical property (*may_treat, may_prevent, has_MoA,* or *has_PE*). As a consequence, on average, the homogeneity score for drug groups is based on less than half of their original single ingredient member drugs.

Acquiring clinical properties for drugs

We extracted clinical properties for 1,701 drugs. However, some drugs do not have properties in all three categories: 1,601 drugs have therapeutic intent properties (*may_treat* or *may_prevent),* 1,612 drugs have mechanism of action properties (*MoA*), and 1,647 drugs have physiologic effect (*PE*) properties.

Computing homogeneity scores

We were able to compute homogeneity scores for 87 of the 90 therapeutic (second level) groups in ATC, with the exception of the groups V01 ALLERGENS, V09 DIAGNOSTIC RADIOPHARMACEUTICALS, and V10 THERAPEUTIC RADIOPHARMACEUTICALS. These groups were either out of scope (V09, V10) or consisted only of drugs without clinical annotations in NDF-RT (V01). For the remaining groups, the number of drugs per group ranges from 126 (in S01 OPHTHALMOLOGICALS) to one (in four groups). As mentioned earlier, we ignored drug groups with only one member drug, reducing the number of groups to 83.

The homogeneity scores range from 1 to 11. Figure 1, top left, shows that the majority (53%) of second level groups in ATC has homogeneity scores of one or two. More precisely, 15 groups (18%) have a score of one, and 29 groups (35%) have a have a score of two. Overall, the ATC classes seem fairly homogeneous for mechanism of action (MoA, Figure 1, middle left) and physiologic effect (PE, Figure 1, bottom left), with a mode of 1 in both cases (distribution of homogeneity scores), and slightly less so for therapeutic intent, where the mode is 2 and the tail longer. *Table 3* shows examples of homogeneous groups in ATC, while *Table 4* illustrates heterogeneous groups.

Comparison to the clinical reference Micromedex

However, in comparison to the groups extracted from Micromedex (MMX), the second level groups in ATC seem to be less homogeneous (see Figure 1, top left and right). 84% of MMX classes show a homogeneity score of 1 or 2, with 57% of all drug classes having a score of 1. As for the ATC groups, we ignored drug classes with only one drug. The discrepancy in terms of homogeneity between second level ATC groups and MMX classes could not be observed for the ATC groups on the fourth level. In terms of both *mechanism of action (MoA)* and *physiologic effect (PE)* annotations, score distributions seem similar to those of the MMX classes (see Figure 1, middle and bottom).

Like all classifications, ATC includes a number of residual groups, designed to accommodate groups not covered by other groups. These groups are likely to be more heterogeneous than regular groups and we recomputed all homogeneity scores after excluding them.
In practice, we ignored the following groups. On the second level we removed all groups of the first level anatomical main group *VARIOUS*, as well as all subgroups containing "other" in their labels. For the fourth level groups, we removed all residual groups, whose codes end with an *X*, e.g., *L02BX Other hormone antagonists and related agents*.
However, we did not observe any significant changes in the distributions of homogeneity scores after excluding these groups, invalidating our hypothesis that they could be significantly more heterogeneous (see Figure 1, right-hand side).

Table 3 - Homogeneous groups in ATC (selected)

Group	Name	# Drugs	Homogeneity score
J01	ANTIBACTERIALS FOR SYSTEMIC USE	92	2
D07	CORTICOSTEROIDS, DERMATOLOGICAL PREPARATIONS	36	1
A10	DRUGS USED IN DIABETES	33	1

Table 4 - Heterogeneous groups in ATC (selected)

Group	Name	# Drugs	Homogeneity score
L01	ANTINEOPLASTIC AGENTS	64	11
B05	BLOOD SUBSTITUTES AND PERFUSION SOLUTIONS	32	8
D03	PREPARATIONS FOR TREATMENT OF WOUNDS AND ULCERS	7	4

Table 5 - Mapping performance by ATC anatomical group

Time	Classification	Drugs in ATC	Single ingredient drugs within scope	Mapped to NDF-RT drug concepts	Mapped to NDF-RT drug concepts with annotations
A	Alimentary tract and metabolism	547	434	275	199
B	Blood and blood forming organs	229	192	108	85
C	Cardiovascular system	558	382	197	166
D	Dermatologicals	345	285	193	168
G	Genito-urinary system and sex hormones	252	183	105	90
H	Systemic hormonal preparations, excluding sex hormones and insulins	75	71	46	34
J	Antiinfectives for systemic use	443	378	203	169
L	Antineoplastic and immuno-modulating agents	251	247	176	120
M	Musculo-skeletal system	205	173	80	62
N	Nervous system	561	482	270	222
P	Antiparasitic products, insecticides and repellents	121	103	40	32
R	Respiratory system	345	281	153	134
S	Sensory organs	267	213	180	164
V	Various	265	140	85	56
	Total	**4464**	**3564**	**2111**	**1701**

Discussion and conclusions

Findings and significance

Because of the warning ATC provides in its documentation about the possible heterogeneity in mechanisms of action and

physiologic effects among drugs in fourth-level groups, we were surprised to find that most groups are actually fairly homogeneous at this level, for both mechanisms of action and physiologic effects.

Moreover, ATC groups are generally not more heterogeneous for mechanisms of action and physiologic effects overall than the MMX classes used as our clinical reference, although ATC groups at the second level are slightly more heterogeneous for therapeutic intent properties.

Finally, our hypothesis that residual classes in ATC would be more heterogeneous than other classes was not verified.

Overall, our findings that ATC classes are generally homogeneous are consistent with the recent adoption of ATC by some researchers for uses outside the realm of pharmacoepidemiology. However, the exact place of ATC in clinical applications remains to be determined.

Limitations

The main limitation of this study is that, as noted earlier, only about half of the ATC drugs can be associated with drug properties in NDF-RT. The two major reasons are that 1) some ATC drugs cannot be mapped to NDF-RT, because they are out of scope or not marketed in the U.S.; and 2) many drugs are not associated with therapeutic intent, mechanism of action or physiologic effect properties in NDF-RT. The incompleteness of NDF-RT in terms of drug properties is the stronger of the two factors, because the exclusion of radiopharmaceuticals and non-marketed drugs is not expected to have any significant impact on clinical applications.

In our comparison of distributions, we did not use any statistical methodology for testing for differences between the distributions. This was on purpose in the context of this exploratory study. Any difference would have been difficult to interpret anyway in the context of an incomplete dataset, as discussed above.

In future work, we plan to explore alternative drug information sources against which ATC could be evaluated. One difficulty, however, is that very few publicly available drug information sources contain reliable clinical information (e.g., DrugBank), most of these information sources being commercial products (e.g., First Databank).

Acknowledgments

References

[1] ATC. Anatomical Therapeutic Chemical (ATC) classification. In; 2012 http://www.whocc.no/atc/.

[2] Jakobsen M, Anker N, Dolleru J, Poulsen PB, Lange P. Study on Drug Costs Associated with Copd Prescription Medicine in Denmark. Clin Respir J 2012.

[3] Adriaenssens N, Coenen S, Versporten A, Muller A, Minalu G, Faes C, et al. European Surveillance of Antimicrobial Consumption (ESAC): outpatient quinolone use in Europe (1997-2009). J Antimicrob. Chemother 2011;66 Suppl 6:vi47-56.

[4] Binder-Foucard F, Reitzer C, Jegu J, Schweitzer B, Koehl F, Kopferschmitt J, et al. Use of psychotropic drugs, systemic antihistamines and medications for cough in 6-year-old children: a survey in the Bas-Rhin Region, France. Pharmacoepidemiol Drug Saf 2012;21(10):1112-7.

[5] Haug JB, Berild D, Walberg M, Reikvam A. Increased antibiotic use in Norwegian hospitals despite a low antibiotic resistance rate. J Antimicrob Chemother 2011;66(11):2643-6.

[6] Neubert A, Hsia Y, de Jong-van den Berg LT, Janhsen K, Glaeske G, Furu K, et al. Comparison of anti-diabetic drug prescribing in children and adolescents in seven European countries. Br J Clin Pharmacol 2011;72(6):969-77.

[7] Chen L, Zeng WM, Cai YD, Feng KY, Chou KC. Predicting Anatomical Therapeutic Chemical (ATC) classification of drugs by integrating chemical-chemical interactions and similarities. PLoS One 2012;7(4):e35254.

[8] Takarabe M, Shigemizu D, Kotera M, Goto S, Kanehisa M. Network-based analysis and characterization of adverse drug-drug interactions. J Chem Inf Model 2011;51(11):2977-85.

[9] Avillach P, Dufour JC, Diallo G, Salvo F, Joubert M, Thiessard F, et al. Design and validation of an automated method to detect known adverse drug reactions in MEDLINE: a contribution from the EU-ADR project. J Am Med Inform Assoc 2012

[10] Pathak J, Murphy SP, Willaert BN, Kremers HM, Yawn BP, Rocca WA, et al. Using RxNorm and NDF-RT to classify medication data extracted from electronic health records: experiences from the Rochester Epidemiology Project.Proc AMIA Annu Fall Symp 2011;2011:1089-98

[11] Pathak J, Weiss LC, Durski MJ, Zhu Q, Freimuth RR, Chute CG. Integrating va's ndf-rt drug terminology with pharmgkb: preliminary results. Pac Symp Biocomput 2012:400-9

[12] Lincoln MJ, Brown SH, Nguyen V, Cromwell T, Carter J, Erlbaum M, et al. U.S. Department of Veterans Affairs Enterprise Reference Terminology strategic overview. Stud Health Technol Inform 2004;107(Pt 1):391-5

[13] NLM. NDR-RT API. In; 2011 http://rxnavdev.nlm.nih.gov/NdfrtAPI.html.

[14] NLM. RxNorm API. In; 2011 http://rxnavdev.nlm.nih.gov/RxNormAPI.html.

Address for correspondence

Olivier Bodenreider -- obodenreider@mail.nih.gov